Towards a Context Model Driven German Geo-Tagging System

Andre Blessing, Reinhard Kuntz and Hinrich Schütze
Institute for Natural Language Processing, Universität Stuttgart
Azenbergstraße 12, 70174 Stuttgart, Germany

ABSTRACT

In this paper, we present a new approach for recognition and grounding of geographic proper names for German. Named Entity Recognition (NER) in German is more difficult than in English because not only proper names, but all nouns start with capital letters, which results in a large pool of potential ambiguous entities. Our approach makes critical use of a geographic knowledge base that is more detailed (down to the level of streets) and more structured than knowledge bases used before. We have designed a three-step model (spotting, typing, referencing) that specifies the sources of information that are necessary for geo-tagging and their dependency relationships. Basic aspects of the model were implemented and evaluated in a proof of concept. The model can be applied to other NER tasks by simply substituting the appropriate knowledge base for the one used here and retraining the model.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: PROBABILITY AND STATISTICS; H.3.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Retrieval; I.2.7 [**ARTIFICIAL INTELLIGENCE**]: NATURAL LANGUAGE PROCESSING

1. INTRODUCTION

Our project is part of the Centre of Excellence "NEXUS - Spatial World Models for Mobile Context-Aware Applications[1]", which is dedicated to research on context models for mobile context-aware systems[8]. A context model typically serves as representation of the real word in digital form. The central focus which is brought about by the aspect of mobility is on object positions. There are several different data representations for the description of spatial information. Communication by means of natural language generally uses symbolic coordinates (e.g. Stuttgart, Germany) to determine spatial information. In contrast to that, spatial information in computational modelling is represented by means of geometric coordinates (e.g. WGS84[2]). To adequately process these heterogeneous data we need mapping procedures to translate between the different viewpoints. In this work we concentrate on the elaboration of the mapping of real-world geo-spatial entities - usually symbolically represented in texts - onto the respective geo-spatial entities objects in the context model (usually represented in a geometric way). In order to realize this, two steps are needed: First we need to identify the entities with possible spatial reading in natural texts, and second we have to connect the referring geo-spatial entity with the recognized typed entity in the context model. We call the first step Named Entity Recognition[16] and the second Geo-Grounding[10].

Context based applications need very detailed and exact information about the location of objects. If you are supposed to meet someone in a foreign city, for example, it will not be enough to mention the name of the city only, but you will need further details on the location, like the names and numbers of streets and places. Current NER systems do not fully meet these requirements because they rarely are able to recognize geo-spatial entities below the level of cities. With our approach we are bridging this gap.

Since ambiguity is a major relevant issue here, it is one of the key challenges of our work to solve it. We differentiate between two main classes of ambiguity: geo/non-geo ambiguity and geo/geo ambiguity [3]. Geo/non-geo ambiguity deals with decisions whether a entity that occurs in our geographic knowledge base (enriched gazetteer) is used in a geographic or non-geographic reading in the current context. E.g.: "Irving Berlin" (no geographic reading possible, because it is a composer's proper name), versus, "In" (town in Palestine). Geo/non-geo ambiguity deals with the mapping of typed entities onto the correct geo-spatial entity in our knowledge base. Examples for this class would be: "Berlin" (capital of Germany or one of the 20 towns in the USA named "Berlin").

A context model offers new previously non-existent possibilities because it contains more fine-grained information. In state-of-the-art work, towns are used as the smallest entities for NER, mostly due to a lack of freely available detailed gazetteers. We will benefit from increasing contexts by additionally incorporating street names, a fact which will generally improve the quality of NER. Consequently we will

[1]funded by the German Research Foundation

[2]World Geodetic System 1984

Yesterday	Peter	spoke	to	Harald.
	SUBJ	V		OBJ
Gestern	sprach	Peter	zu	Harald.
	V	SUBJ		OBJ

Table 1: Examples for free word order in German

use a topology including the following toponyms[3]: territory, region, city, street.

NER research up to now has mainly been concerned with English. But it is necessary to take into consideration the special features of languages other than English. In this paper, we propose a design for a German geo-tagging system. Many aspects of current models for English NER do not apply to German because of the different linguistic structure of German. In German, not only proper names, but all nouns start with capital letters and therefore the amount of potential proper names increases dramatically and so does the number of cases which show ambiguity between geo-spatial proper names and normal nouns. In addition to that, the local contexts have more variance than the English ones (rigid syntax: subject-verb-object) because German is a language with relatively free word order. If in English a capitalized word in subject position occurs before a communication verb, the word must be a proper name (person or organization, see figure 1). This principle does not apply to German, since the finite verb can take three different positions in the sentence (first, second, last), whereas the subject need not necessarily adopt the front position.

Named Entity Recognition is a very important part of Information Retrieval and is quite popular since the Message Understanding Conferences (MUC) [7] in the late 1980s and 1990s. At MUC-7 most systems for NER were rule-based and achieved good result, but setting these systems up is expensive. Moreover they work only on special domains, adaptation to other domains is labour-intensive. To solve this disadvantage a number of methods from machine learning [6] have many applied in the field of NER in last 10 years.

In the next section we will present related recent work. In section 3 our approach will be explained in detail. Results of the first partial implementations will be evaluated in section 4. Finally we will conclude and provide some ideas for future work.

2. RELATED WORK

In the area of NER and geo-tagging, or, geo-grounding, some research has already been done. In the following we will present some work which is related to our topic and serves as the basis for our own studies. The majority of this work is based exclusively on English NER systems, which is the reason why many of these methods cannot be straightforwardly transferred to German NER.

Leidner [10] describes the standard geo-grounding approach. The main focus of his work is on geo-geo disambiguation, where all possible references are drawn in a map and those entities which span a polygon with the smallest area are chosen. In our work we are using this aspect of minimal distances for disambiguation as well, but we complement it with further aspects (e.g. part-of relations: Bahnhofstraße is part of Berlin and Berlin is part of Germany).

In [3], Web-a-Where, a fast and efficient algorithm for geotagging on Web pages, is presented. Web-a-Where not only references geographical entities, it also assigns to each web page a geographic focus. Web-a-Where geo/non-geo and geo-geo disambiguation. We cannot use the methods implemented in Web-a-Where because German does not satisfy many of the basic assumptions that hold for English. We discus this in more detail in section 3.1.

In [12] it is shown that NER can work successfully without lists (as they can be derived from knowledge bases and gazetteers). This is mainly applicable to the recognition of names of persons and organisations, because in these cases there are more simple indicator patterns: title (Mr., Mrs.), multi-word-expressions (name surname), and abbreviations for business and commerce (inc., plc.). However, the authors show for geographic NER that results with gazetteers are significantly better than results without any form of knowledge base. Therefore we see our enriched knowledge base as an important building block in the architecture of geo-tagging systems.

In 2003, a CONLL Shared Task for language-independent NER took place([14]). The data were English and German [4]. It was shown that all results for German were at least 5 percent worse than results for English. This supports our assertion that German NER is harder and needs a special approach. The best evaluation results were achieved by machine learning methods ([6],[9],[5]).

In [13] and [15] several approaches for German NER system were introduced. These system have no special focus on geo-spatial entities, they only use the standard categories from MUC (person, organisation, location.). Their approaches are too general for our task and include no geo-grounding aspect.

3. FORMAL DESCRIPTION

In this section our approach is formally described. First, we explain the goals of our model, then we sketch the necessary preprocessing steps and finally we will describe the formal definition of our model.

3.1 Goals

Our goal is to achieve the highest possible recall, but at the same time maintain high precision. In previous work, this kind of trade-off was not well balanced because entities which in most of the cases had no geo-spatial meaning were neglected and therefore not recognizable if used in their geo-spatial meaning. This selection was in most cases achieved by a twofold procedure: locations with small impact (low number of inhabitants), and entities which frequently appear in lower case (in German this is not feasible because all nouns are written with initial capitalization) were removed from the gazetteers that were used in these experiments.

[3]Definition of toponyms: all topographical objects named with a proper name. It includes plane, linear and point objects.

[4]Articles of Frankfurter Rundschau, years 1993-1994

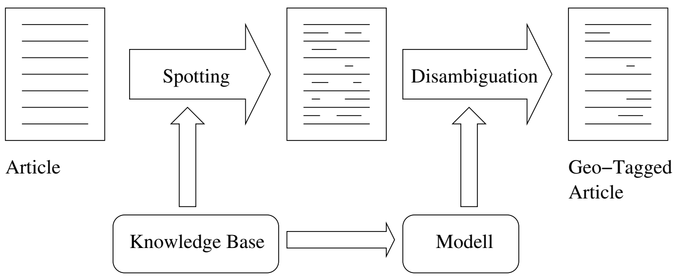

Figure 1: Complete Geo-Tagging Processing Chain

Our goal is to be able to recognize all geo-spatial entities, including rare ones. While one might argue that excluding rare geo-spatial entities on the city level is appropriate, it is problematic on more detailed levels.

We want to recognize geo-spatial entities (including rare ones) using features which are derived from the current context. A further challenge arises because of geo-spatial entities which are multi word expressions, especially if they include a preposition and therefore could easily be confounded with a prepositional phrase. Our most important aim is to develop a model apt for general usage where only the knowledge base must be exchanged or adapted. In this work we developed this model and used a knowledge base which was derived from the NEXUS context model. Thus our model should easily be applicable to other tasks if the knowledge base is exchanged.

3.2 Work-flow

Figure 1 depicts the work-flow of our geo-tagging process. The first step is spotting: each entity which is in our knowledge base is marked. Many of the marked entities have several types, so a disambiguation process is needed. In section 3.5.3 we describe this process in detail. Our system is designed as a knowledge-based system. The system is easily adaptable to other tasks (e.g. a detailed person NER system) if the knowledge base is substituted.

3.3 Knowledge Base

We are using two main sources for our knowledge base, both are parts of the NEXUS context model. A standard gazetteer [1] is used to get a global geo-spatial knowledge for the world, but it is not fine-grained enough for our purposes. From a data provider for navigation systems [2] we obtained a detailed geographic knowledge base for all of Germany. This data set represents all streets, cities, regions and states inside Germany. To use this resource as a knowledge base we had to do a few preprocessing steps.

- Normalisation: Replacement of all non-alphabetic characters with spaces
- Capitalisation: Adjustment of lower and upper case
- Expansion: Generation of abbreviations and writing variants (Boston expand also to Boston, MA)
- Extension: Inclusion of misspelled variants (Schanbacherstr. instead of Schanbacher Str.)

3.4 Corpora

Our experiments are run on two different text corpora. The Stuttgart Newspaper Corpus can be seen as a reference corpus comparable to corpora used in recent studies in NER. Thereby we can compare our results with other related work. For the first time in a geo-tagging research project, we use the German Wikipedia. In Wikipedia, no uniform style of writing is used because many different authors are responsible for the articles. That is why some of the texts contain different variants of wording or even grammatical mistakes, which complicate natural language processing. In this we see a big challenge to test the robustness of our methods and to achieve the highest possible coverage on different domains and styles. Moreover the amount of available information via Wikipedia is still increasing and very up to date. So it is an excellent resource for getting highly relevant information for context-aware systems.

3.5 Model

We use the following notation for the specification of our model:

$\mathbb{E}$	set of all possible entities	
E	current sequence of entities (from spotter)	$E \subset \mathbb{E}$
e_i	current entity	$e_i \in E$
$\mathbb{T}$	set of typed entities (incl. non-geo)	
T	current sequence of typed entities	$T \subset \mathbb{T}$
t_i	typed entity for e_i	$t_i \in T$
$\mathbb{G}$	set of geo-spatial objects in database	
G	current sequence of geo-spatial objects	$G \subset \mathbb{G}$
g_i	geo-spatial object for t_i	$g_i \in G$
K_{ci}	context knowledge available at pos. i	
K_g	set of global (world) knowledge	
L_c	linguistic context	

We distinguish the following types of geo-spatial objects:

$\mathbb{G}_s$: street(street) {A1, ..., Zypressenweg}

$\mathbb{G}_c$: city{Aachen, ...,Zwickau}

$\mathbb{G}_r$: region{Baden, ..., Wuerttemberg}

$\mathbb{G}_t$: territory {Afghanistan, ..., Zypern}

$\mathbb{G}_n$: non-geo{non-geo-spatial reading}

3.5.1 *Mapping Process*

The spotting process returns a sequence E of entities $e_1, ..., e_n$. As shown in Figure 2, we then want to type (Text → Typed entities) and reference (Types Entities → Map) each of these entities. The leftmost box contains the input text with marked entities e_i. Each entity can assume one or more of the

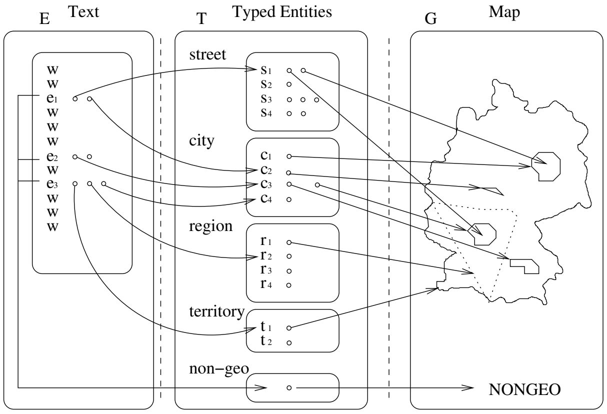

Figure 2: Mapping Processing

available five NER types (street, city, region, territory, non-geo). Example: Berlin-city, Berlin-non-geo (the composer), Berlin-region (federal state in Germany). A typed entity has in turn several mapping possibilities to a geo-spatial entity in the context model (Berlin-city → Berlin OH, Berlin PA etc.). Our goal is to find a mapping function which assigns the current entity e_i the best typed entity t_i and then assigns t_i the best geo-spatial entity g_i, given the current context.

For the type non-geo, we introduce the dummy object NONGEO. All entities that were typed as non-geo are mapped to NONGEO.

3.5.2 Knowledge Sets

K_{ci} is the set of pairs <entity, geo-spatial object > that has been explicitly mentioned up to position $i-1$. This set is constantly increasing during geo-tagging because each geo-tagged entity is included.

In addition, when processing text that is associated with a particular locality, there is a set of geo-spatial entities associated with the locality that is immediately available for geo-tagging without having been mentioned explicitly. For example, if a street name in an article of Stuttgarter Zeitung is noted without any city references, we can infer that the current city in this context must be Stuttgart. We initialize K_{c1} with the set of prominent geo-spatial entities associated with the locality of publication.

There is a second set K_g that contains a set of universally prominent geo-spatial entities such as 5th Avenue, Champs Elysees or New York City. These entities are always capable of being referenced directly – without prior geographic context. Non-prominent entities (e.g., a small town like Backnang in Germany) require some form of geographic context ("Backnang in Germany", "we took the train to Stuttgart and then continued on to Backnang" etc). In contrast, prominent entities can be used without such geographic context. We distinguish K_{ci} and K_g because entities in the former are available for grounding (the default interpretation for "Bahnhofsstraße" in the Stuttgarter Zeitung is the corresponding street in Stuttgart, a member of K_c), but not for the latter ("Peter Cooper Rd" cannot be used without geographic context even though New York City is in K_g; New York City has to be mentioned explicitly and then added to K_{ci}).

3.5.3 Calculus

Our goal is to find the best sequence G of geo-spatial objects given an input sequence E of entities, a linguistic context L_c, and contexts K_{c1} and K_g:

$$G_{best}(E) = arg\max_{G} P(G|EL_cK_{c1}K_g)$$

We define an intermediate step to simplify the two different disambiguation steps. In the intermediate step we assign each entity to a typed entity:

$$\begin{array}{cccc} g_1 & g_2 & \dots & g_n \\ | & | & & | \\ t_1 & t_2 & \dots & t_n \\ | & | & & | \\ e_1 & e_2 & \dots & e_n \end{array}$$

For example, "Berlin" might be typed as city and "main street" as street.

We assume that t_i only depends on e_i, and the linguistic and geographic contexts:

$$P(t_i|E, L_c, K_{c1}, K_g) = P(t_i|e_i, L_c, K_{ci}, K_g)$$

Recall that K_{ci} contains all previously made geo-tagging decisions in the article. That is, once e_i has been assigned to g_i, we add $< g_i, t_i >$ to $K_{c(i+1)}$.

We assume that g_i only depends on t_i, and the geographic contexts:

$$P(g_i|t_i, E, L_c, K_{c1}, K_g) = P(g_i|t_i, K_{ci}, K_g)$$

Assuming conditional independence of the typing and referencing decisions in the n positions of the sequence, we obtain:

$$
\begin{aligned}
G_{best}(E) &= arg\max_G P(G|E, L_c, K_{c1}, K_g) \\
&= arg\max_G \prod_{i=1\dots n} P(g_i, t_i|E, L_c, K_{ci}, K_g) \\
&= arg\max_G \prod_{i=1\dots n} P(g_i|t_i, E, L_c, K_{c1}, K_g)P(t_i|E, L_c, K_{c1}, K_g) \\
&= arg\max_G \prod_{i=1\dots n} P(g_i|t_i, K_{ci}, K_g)P(t_i|e_i, L_c, K_{ci}, K_g)
\end{aligned}
$$

The parameters $P(g_i|t_i, K_{ci}, K_g)$ and $P(t_i|e_i, L_c, K_{ci}, K_g)$ are partially set by constraints we impose on the model and partially estimated from annotated training data.

Constraint 1. One sense per article. If an entity e_i has been tagged as g_i, then all subsequent tokens of the same entity also have to be tagged as g_i. For example, if a sentence mentioning Berlin, the capital of Germany, is followed by the expression "Berlin is cool.", then "Berlin" cannot refer to "Irving Berlin." We formalize this constraint as follows:

$$P(g_i|t_i, K_g, K_{ci}) = 1.0 \quad \text{if} \quad < g_i, t_i > \in K_{ci}$$

Constraint 2. Completeness of knowledge base. If our geographic knowledge base does not contain a typed entity, then it does not exist:

$$P(g_i|t_i, K_g, K_{ci}) = 0.0 \quad \text{if} \quad < g_i, t_i > \notin \mathbb{G}$$

A special case of this constraint is that an entity that was typed as non-geo can only be assigned to the dummy object NONGEO.

Constraint 3. Mandatory geographic context. A geo-spatial object can only be introduced if its geographic context is accessible. There are three types of accessibility: (i) The object is in K_{ci}. (ii) The object is in K_g. (iii) The object is an immediate child of an object in K_{ci}.

To formalize this notion, we define a part-of relation. It obeys the following axioms:

$$\forall g_s \in \mathbb{G}_s \exists g_c \in \mathbb{G}_c \forall g'_c \in \mathbb{G}_c : g_c = g'_c \Leftrightarrow R_{partof}(g_s, g'_c)$$
$$\forall g_c \in \mathbb{G}_c \exists g_r \in \mathbb{G}_r \forall g'_r \in \mathbb{G}_r : g_r = g'_r \Leftrightarrow R_{partof}(g_c, g'_r)$$
$$\forall g_r \in \mathbb{G}_r \exists g_t \in \mathbb{G}_t \forall g'_t \in \mathbb{G}_t : g_t = g'_t \Leftrightarrow R_{partof}(g_r, g'_t)$$

That is, each street is part of exactly one city, each city is part of exactly one region, and reach region is part of exactly one territory.

With this part-of relation, we can formalize Constraint 3 as follows:

$$
\begin{aligned}
P(g_i|t_i, K_g, K_{ci}) \;&=\; 0.0 \\
\text{if} \quad \neg \quad (\quad & < g_i, t_i > \in K_g \\
& \vee < g_i, t_i > \in K_{ci} \\
& \vee \exists g_j, t_j [R_{partof}(g_i, g_j) \wedge < g_j, t_j > \in K_{ci}] \\
)
\end{aligned}
$$

In preliminary experiments (see below), we have found that imposing these constraints very rarely causes tagging errors. However, it is important to build a good linguistic preprocessing module. For example, the spotter must recognize "Jack London" and "Berlin OH" as entities otherwise phrases like "from London to Jack London's home town" or "from Berlin, Germany, to Berlin OH" cannot be tagged correctly. Also, without proper linguistic preprocessing different forms of the same name cannot be recognized (e.g., the geographic database may list "Freiburg im Breisgau", but this city is often just referred to as "Freiburg").

Parameters $P(g_i|t_i, K_g, K_{ci})$ that are not covered by the constraints and the parameters $P(t_i|e_i, L_c, K_{ci}, K_g)$ are estimated from a labeled training set (initial work on this estimation problem is discussed below). In the training set, each instance of typing and each instance of tagging are represented as labeled feature vectors. We use external and internal evidence to build feature vectors. External evidence includes the right and left linguistic contexts of the current entity (window size 3); and whether the entity already occurs with the same type / object or a different type / object in K_c or K_g. We extract words and bigrams from the linguistic contexts and add them to the feature vectors. As internal evidence we use the shape of the entity (e.g., does it contain a hyphen), its suffix, its capitalization and a shortened representation (conflating capitalised characters, conflating non-capitalised characters, conflating special characters).

We use SVMs to build a probabilistic classifier from the training set. SVM scores are then transformed into probabilities using logistic regression.

3.5.4 Enumerations

Entity enumerations (e.g. coordination) in texts have a special status in our model because all elements of an enumeration are mapped to the same geo-spatial type. We have confirmed this hypothesis in a small empirical study (we automatically extracted all enumerations of our corpora and found no exceptions). Therefore, for geo-tagging enumerations, all sequences of entities $e_i, ..., e_{i+n}$ which are separated by one of the tokens {, ; : - and or} (and no other tokens) are combined to a new entity for this we find the right NER type. After that all compound entities are assigned to the typed entities in the same type and the grounding process is run separately for each.

4. RESULTS

The whole system has not yet been fully implemented. Thus far, we have implemented the street name classifier and the city name classifier.

4.1 Street Name Classifier

We randomly selected 15,000 entities for street names from two corpora (German version of Wikipedia and the daily newspaper Stuttgarter Zeitung). Each of the entities was labelled manually by an annotator who based her classification on the entire sentence context. The annotated data show a large inter-annotator-agreement (kappa value above 0.8 [4]). We divided them into three sets: training, testing, and evaluation set. We chose SVM as our classification method and converted each entity into a feature representation. By adopting this relatively simple approach, our method achieved an F_1 score [11] of 90% (baseline was below 80%).

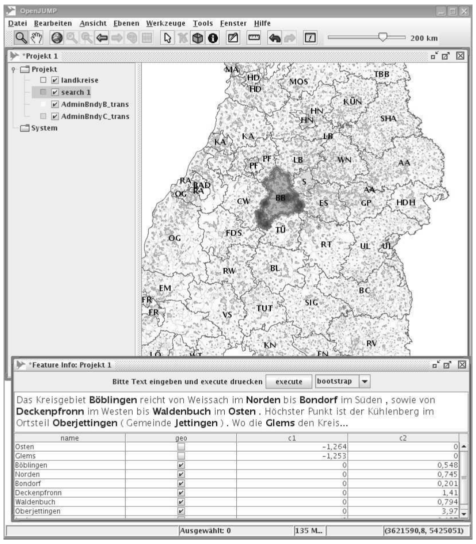

Figure 3: Prototype for a City Name Classifier

We expect that we will substantially improve on this result when we fully implement the model. Additionally, we want to test how much one could minimize the cost of annotation to lower the cost of building a complete geo-tagging system.

4.2 City Name Classifier

We also developed a prototype for NER of location names (see Figure 3 for a screenshot of the prototype). Here again we reached an F_1 score of 90%. The baseline is 75% (distribution of geo vs. non-geo usage in the evaluation set). We tested a number of methods for classifying entities as geo or non-geo. The supervised approach achieved the best results, but an unsupervised approach also achieved a substantial improvement of 86% over the baseline of 75%.

5. CONCLUSION

In this paper, we have presented a new approach for geo-tagging that is applicable to languages like German that pose a more difficult geo-tagging problem than English. Our approach makes critical use of a geographic knowledge base that is more detailed (down to the level of streets) and more structured than knowledge bases used before. The model can be applied to other NER tasks by substituting an appropriate knowledge base for the one used here and retraining the model. We have designed a three-step model (spotting, typing, referencing) that specifies the sources of information that are necessary for geo-tagging and their dependency relationships. Basic aspects of the model were implemented and evaluated in a proof of concept.

6. FUTURE WORK

The most important next step in our research is to fully implement and evaluate our model. To minimize annotation costs when creating training sets, we plan to use semi-supervised methods like bootstrapping and active learning. The representation of contexts is critical for high accuracy in geo-tagging. Here, we plan to explore feature selection, feature weighting, linguistically based features (part of speech tags, partial parses), larger context windows, and different ways of representing K_c and K_g.

7. REFERENCES

[1] Gazaetteer, http://www.bevoelkerungsstatistik.de.

[2] Navteq, http://www.navteq.com.

[3] E. Amitay, N. Har'El, R. Sivan, and A. Soffer. Web-a-where: geotagging web content. In *SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 273–280, New York, NY, USA, 2004. ACM Press.

[4] J. Carletta. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2), 1996.

[5] H. L. Chieu and H. T. Ng. Named entity recognition with a maximum entropy approach. In *Proceedings of CoNLL-2003*, 2003.

[6] R. Florian, A. Ittycheriah, H. Jing, and T. Zhang. Named entity recognition through classifier combination. In *Proceedings of CoNLL-2003*, 2003.

[7] R. Grishman and B. Sundheim. Message understanding conference - 6: A brief history. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, 1996.

[8] M. Grossmann, M. Bauer, N. Hönle, U.-P. Käppeler, D. Nicklas, and T. Schwarz. Efficiently managing context information for large-scale scenarios. In *Proceedings of the 3rd IEEE International Conference on Pervasive Computing and Communications*. IEEE Computer Society, 2005.

[9] D. Klein, J. Smarr, H. Nguyen, and C. D. Manning. Named entity recognition with character-level models. In *Proceedings of CoNLL-2003*, 2003.

[10] J. L. Leidner, G. Sinclair, and B. Webber. Grounding spatial named entities for information extraction and question answering. In *Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references*, pages 31–38, Morristown, NJ, USA, 2003. Association for Computational Linguistics.

[11] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[12] A. Mikheev, M. Moens, and C. Grover. Named entity recognition without gazetteers. In *EACL*, 1999.

[13] M. Rössler. Corpus-based learning of lexical resources for german named entity recognition. In *In Proceedings of LREC*, 2004.

[14] E. F. Tjong Kim Sang and F. De Meulder. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In W. Daelemans and M. Osborne, editors, *Proceedings of CoNLL-2003*. Edmonton, Canada, 2003.

[15] M. Volk and S. Clematide. Learn - filter - apply - forget. mixed approaches to named entity recognition. In *NLDB*, pages 153–163, 2001.

[16] G. Zhou and J. Su. Named entity recognition using an hmm-based chunk tagger. In *ACL '02*, Morristown, NJ, 2001.